

(11)Publication number : 05-290549  
(43)Date of publication of application : 05.11.1993

(51)Int.Cl.

G11B 27/034  
H04N 5/76  
H04N 5/782

(21)Application number : 04-112404

(71)Applicant : NIPPON TELEGR & TELEPH CORP  
 <NTT>

(22)Date of filing : 06.04.1992

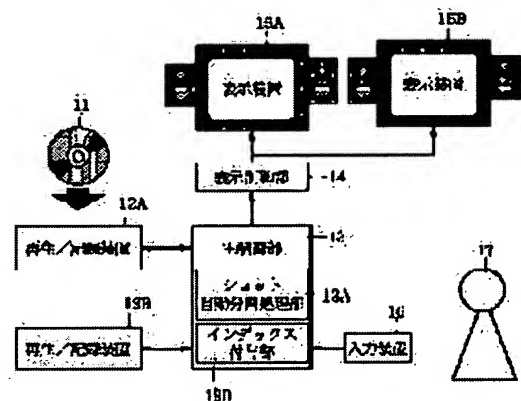
(72)Inventor : HASEYAMA HIROSHI  
OBA YUJI  
TANAKA TOMOAKI

**(54) VIDEO EDIT SUPPORTING SYSTEM**

**(57)Abstract:**

**PURPOSE:** To make an automatic division of a shot by detecting a plurality of cut points based on a plurality of division references in editing video.

**CONSTITUTION:** A video source 11 is reproduced by a reproducing/recording device 12A and displayed on a display device 15. An operator 17 makes an input from an input unit 16 and gives an index at an index supply section 13B. At the time of editing, a shot automatic division processing section 13A makes a choice from among a plurality of cut points divided based on the given references to be displayed on a display device 15B. When simple directions changing the shot length to 'long' and to 'short' are supplied from an input unit 16, another cut point is displayed on the display device 15B. By repeating such processing, the video editing can be performed.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

**[Date of final disposal for application]**

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

**Complete English Translation of JP 05-290549****[SCOPE OF CLAIMS FOR PATENT]****[Claim 1]**

A video edit supporting system serving to support an apparatus for editing video and audio characterized by comprising:

an index supply section for recording index information serving as a marker for a content of the video and audio;

a shot automatic division processing section for automatically detecting shots based on a plurality of references previously provided and showing, at the time of editing, a cut point defined based on another reference as a candidate with respect to an instruction to change a length of the shot defined based on a particular reference; and

a display device of displaying a candidate shot.

**[Claim 2]**

A video edit supporting system as claimed in Claim 1, characterized in that the shot automatic division processing section employs audio information as one of the plurality of references used for the automatic detection of the shot.

**[DETAILED DESCRIPTION OF THE INVENTION]****[0001]****[FIELD OF INDUSTRIAL APPLICATION]**

The present invention relates to a video edit supporting system for automatically editing videos in response to an instruction given thereto in editing the stored videos. More specifically, a sequence of videos (hereinafter, referred to as shot) having a certain meaning is automatically divided from a sequence of stored videos (hereinafter, referred to as video file) and displayed. A user, watching the display, provides a simple instruction such as "longer" or "shorter" to thereby change a length of the shot to be automatically divided and displays the shot again. The present invention relates to the video edit supporting system for performing the editing work by repeating the foregoing process.

[0002]

[PRIOR ART]

Conventionally, a shot division was performed referring to a memorandum relating to video contents after the entire video file was carefully watched and listened in editing the stored videos. Further, in a method of dividing the shot confirming the video contents in a still picture state, a heavy error is generated between an initially conceived shot-division image and a post-editing image by incorrectly recognizing where the shot is consecutive, or the like. In order to prevent the problem from happening, it was necessary to repeatedly confirm the post-editing video and the linkage of the audio in performing the editing by, for example, repeatedly copying a necessary part

after a sequence of moving pictures is reproduced and confirmed in a frame feeding state. There was also a disadvantage that it was not possible to accurately perform the shot division in the case of speedy editing and in terms of an editing image conceived while watching the consecutive videos. Further, it was not possible to automate an entire process relating to the shot division, which therefore required a manual work.

[0003]

The video, which is rather difficult to be provided with a search key, cannot be easily subjected to a search processing in contrast to a letter. However, the search of the moving picture employs a method of previously appending a signal recognizable by a human as a marker, that is an index, to a specific frame to thereby perform the search referring to the index, or the like. In a videotape recorder of a certain type, as an example, a search signal is recorded on the videotape in a corresponding manner when recording starts to thereby more efficiently search a starting point of the recording, and there is a laser disk adapted in the same manner provided with a function such as a chapter search. The index can be mechanically appended to a block of specific videos. However, when the indexing is required beyond the level of a simple automatic recording such as searching the starting position of the recording, it had to be done by confirming the content of each still picture, which had no choice but to depend on a manual labor.

[0004]

Methods of detecting the cut points are recited in Nagasaka, Tanaka: 1Q-5, pp. 642 of the National Convention of Information Processing Society of Japan, first half of 1990, Otsuji, Sotomura, Oba: D-675, pp. 7-387 of the Spring National Convention of Institute of Electronics 1991, Otsuji, Sotomura, Oba: D-210, pp. 6-212 of the Autumn National Convention of Institute of Electronics 1991, and the like.

[0005]

Further, as an example of the shot division references, a method of extracting an object area focusing on a photogenic object in the video includes methods recited in Nagao, Soma, Ando et al.: IE90-105 of the Technical Report of Institute of Electronics 1990, Akyoin, Ide, Nagahashi: D-267, pp. 6-269 of the Autumn National Convention of Institute of Electronics 1991, Inui, Hosono, Yamada: D-191, pp. 6-193 of the Autumn National Convention of Institute of Electronics 1991, and the like.

[0006]

As a post-area-extraction processing can be mentioned methods recited in Ueda, Mase, Suenaga: D-220, pp. 6-222 of the Autumn National Convention of Institute of Electronics 1991, Inoue, Koyama: D-200, pp. 6-202 of the Autumn National Convention of Institute of Electronics 1991, Koyama, Inoue: 21-1, pp. 381 of the Annual convention of TV Academy 1991, ITEC '91, and the like.

[0007]

[PROBLEMS TO BE SOLVED BY THE INVENTION]

In order to perform the shot division in editing the stored videos, it is necessary to carefully watch and listen the entire video file, which is time-consuming. Further, to perform the shot division confirming the video contents in the still picture state included a disadvantage that the heavy error was generated between the initial editing image conceived when the videos were consecutively displayed and the post-editing image. Therefore, in order to eliminate the disadvantage, it was necessary for the videos, which had been edited then, to be repeatedly confirmed while simultaneously performing the editing work by means of the still picture. This was disadvantageously quite complicated and time-consuming.

[0008]

It is rather difficult to append the search key to the video. The indexing beyond the level of the simple automatic recording such as searching the starting position of the recording always required a manual work. Therefore, it was very difficult to record an image captured at the moment when it was viewed in a corresponding manner with respect to a relevant video frame.

[0009]

Therefore, a main object of the present invention is to provide a video edit supporting system capable of detecting cut

points and further automatically dividing a shot.

[0010]

[MEANS OF SOLVING PROBLEMS]

A video edit supporting system according to the present invention comprises an index supply section for recording index information serving as a marker for a content of the video and audio, a shot automatic division processing section for automatically detecting shots based on a plurality of references previously provided and showing, at the time of editing, a cut point defined based on another reference as a candidate with respect to an instruction to change a length of the shot defined based on a particular reference, and a display device of displaying a candidate shot. The shot automatic division processing section employs audio information as one of the plurality of references used for the automatic detection of the shot.

[0011]

[ACTION]

According to the present invention, a plurality of shot divisions ranged from a short shot length division to a long shot length division is automatically performed in advance based on a plurality of division references. A user is provided with a division example similar to a shot division of a sequence of video frames having identical index information for any part indexed by the user and a division example at an intermediate



level between the long-division-length division and the short-division-length division for any unindexed part. The user repeatedly communicates a simple message such as "longer" or "shorter" in comparison to his/her own image for the shot division videos to thereby perform the video editing. In other words, because breaks in the videos are automatically detected based on the plurality of conditions (references), after the content of the video file is browsed once, the editing work only requires such processes as issuing the simple instruction to the videos per shot displayed based on the index information set by the user at that time, determining an arrangement order of the shots, and combining the videos with the contents of a plurality of video files when necessary. In the foregoing manner, time consumed for such a complicated work as recording video frame numbers and the corresponding video contents, recording a shot division range conceived by the user, and the imaging of the video contents can be saved, thereby efficiently performing the editing work. Further, any erroneous editing caused by wrong memory and false recognition can be thereby avoided.

[0012]

With respect to video information, in addition to such a processing as extracting the object area and pursuing the object after the area extraction, characteristics of the respective area-extracted objects are extracted. The characteristics of

the respective objects are extracted based on conditions such as depth of focus, moving status, changing state, and relationship with respect to camera operation, which are referenced for the shot division. As an example of the shot division, while one or a plurality of extracted specific objects remains in a screen, it is regarded as one shot. Further, the videos between scene changes can be also divided as one shot if a specific object disappears from the screen.

[0013]

Regarding audio information, breaks of music, vocalization status of a particular speaker by means of speech recognition, noise level, noise property, and the like are employed as the division references.

[0014]

[EMBODIMENTS]

Fig. 1 is a diagram illustrating a configuration according to an embodiment of the present invention. Reference symbols in the drawing respectively denote the followings. 11 denotes a video source, which is a video medium recorded on a laser disk, videotape, or the like. The video source 11 is fetched into a main control section 13 by means of a reproducing/recording device 12A. 13B denotes an index supply section for indexing the video contents through a key operation conducted by a user 17. When the user 17 operates an optional key watching the video contents, the index supply section 13B records a video frame

number and the detail of the key. 13A denotes a shot automatic division processing section, wherein a plurality of shot divisions are performed to one video file and memorized. 12B denotes a reproducing/recording device, wherein the videos per shot are recorded on the video medium when the shot division desired by the user 17 is completed based on an interaction with the user 17. In the case where the videos per shot are rearranged in a different order or combined with any shot of other video files at a final step of the editing work, the reproducing/recording devices 12A and 12B are alternately used to thereby interchange the videos. 14 denotes a display control section, which is a video synthesis control section for displaying the shot division videos and an output of the main control section 13. 15A and 15B are display devices for displaying the contents synthesized by the display control section 14. 16 denotes an input unit such as a key board and mouse, wherein the user 17 gives the instruction to correct the index information and shot division range.

[0015]

Fig. 2 shows examples of the cut point detection and shot division. 21-1 - 21-4 show an example of the division per scene change as an example of the cut point detection. More specifically, shooting is temporarily halted at respective intervals between 21-1, 21-2, 21-3, and 21-4 and restarted. Therefore, the videos represent scene changes which are quite

different from the respective previous videos and detected as the breaks of the cut points. 22-1 and 22-2 are divided based on the index information set by the user 17. It is seen that 22-1 focuses on a plane, while 22-2 focuses on an automobile. However, at a break between 22-1 and 22-2, a video of the automobile is falsely set as a video of the plane. This shows that a lag is generated between timings when the key is pressed and released because the setting was done watching the videos at a normal speed. 23 is an example of the division based on the audio information, in which a continuous music including no break is recorded on the entire videos. Thus, one video file can be divided in a plurality of manners depending on which division method is employed.

[0016]

Fig. 3 is an illustration of examples of the automatic shot division and an operation at the time of adjustment. Reference numeral 31 denotes a video file. Reference numeral 32 denotes an index set by the user 17, in which index letters A and B are used. It is assumed that the plane and automobile are the video contents in the index letters A and B. Reference numeral 33 denotes pre-adjustment division information for display, and 34 to 38 denote a plurality of examples of the shot automatic division. Looking at a display status prior to the adjustment, the division based on the index information is employed where the index 32 is set, while the division 36

performing the intermediate division is employed where the index 32 is not included.

[0017]

Next, a processing in the case of changing a division point in accordance with the user 17's instruction is described referring to Fig. 4. Four typical methods are described here, and reference numerals 40 to 43 show the methods 1 to 4, respectively. 40-8 denotes a part to be displayed prior to the adjustment, which corresponds to 33 in Fig. 33. 40-9 denotes a current division point requesting the user 17's judgment. If the user gives the instruction "longer", a next candidate is selected from a plurality of division points based on the different division references temporally present prior to the next division point 40-11 of the division method to be displayed prior to the adjustment and after the current division point, that is a range of selection objects.

[0018]

In the method 1, the division point 40-10 closest to a value representing half of a time interval between the current division point and the next division point is selected. The video of the division point is displayed to the user 17, and then the next instruction is received. In the method 2, the temporally earliest division point is selected from the division points to be selected. In the method 3, the farthest division point is selected. In the method 4, the division point is selected

by means of a condition expression  $n/\epsilon$  for a typical search method used in the case of searching relevant data from a plurality of data. Here,  $\epsilon$  is 2.718 as the base of natural logarithm, and  $n$  is 1 or 2.

[0019]

Further, though not described here, a range beyond the next division point may be selected as the selection object, or, if there is a tendency of regarding the break of the music, for example, as a division point in accordance with the user 17's preference for extracting the division point, division information on the display object thereafter may be replaced by the division based on a current reference or reference in compliance with the tendency, thereby constituting a more user-friendly system.

[0020]

Fig. 5 shows an example of screens of the display devices. Reference numerals 50 and 60 respectively correspond to the display device 15A and 15B. To describe the devices following operation steps, the user 17 previously looks over the video contents using a moving picture window 62 and a video operation button 63. Further, the keyboard shown in Fig. 1 is used to set the index information with respect to the video contents. The automatic division starts and ends in response to instructions from buttons 64-1 and 64-2. When the button 64-1 for starting the division is operated, the cut point detection

and shot division are started based on the different kinds of video division references. When the divisions of a plurality of kinds are completed, videos 52-1 to 52-3 for an automatically divided shot are displayed, wherein the video 52-1 shows a leading video of the shot, the video 52-2 shows an intermediate video of the shot, and the video 52-3 shows a final video of the shot. Reference numeral 51 denotes a still picture window for displaying two types of contents, wherein a leading video of a shot before the video 52-1 or a shot of the shortest division between 52-1 and 52-2. In the same manner, a window 53 displays a still picture after the video 52-3 or between the videos 52-2 and 52-3. Thus, the system can be used demanding no confirmation of the video contents.

[0021]

A shot division point is extended or shortened by means of buttons for "longer" and "shorter" shown with reference numerals 54 and 56. A part of the shot where the shot division point is extended or shortened is determined by clicking one of the frames of the videos 52-1 and 52-3. When buttons 55 or 57 respectively for "longer in altogether" and "shorter in altogether" are pressed, the cut division points for both the leading and final videos are extended or shortened.

[0022]

When the videos of the shot agree with the image conceived by the user 17, a button 58 for "adjustment end" is pressed.

Thereby, the relevant shot is set in a display window 61 for displaying a post-adjustment shot. In this case, the contents of the shot displayed are, as in the previous case, initial, intermediate, and final videos. When all of the shots are adjusted or the adjustment is ceased by means of the button for ending the division along the way, each shot is now edited. A method of editing per shot includes an order rearrangement, copying, deletion, incorporating a shot from other video files using a button 64-3. Further, in the case where the videos displayed on the respective windows are indexed, the videos can be synthetically displayed to thereby confirm the initial division image conceived by the user 17. The instructions of "longer" and "shorter" can be arranged to be audio instructions through a microphone, which significantly simplifies the operation.

[0023]

Fig. 6 is a flow chart for describing an operation of the main control section, wherein (S1) to (S12) show respective operation steps. When the editing work is started and the user 17 first confirms the video contents (S1), the main control section 13 monitors a keyboard operation (S2). When a key is pressed, the main control section 13 memorizes the relevant frame number and key content as the index information (S3). The keyboard operation is continuously monitored when no instruction to start the division is issued (S4). When the video contents



are confirmed and the instruction to start the division is given, the automatic division processing based on the plurality of division references and the plurality of division ranges are memorized (S5). When all of the different kinds of automatic divisions are completed, the display starts in accordance with the pre-adjustment division information 33 for display, which was described referring to Fig. 3 (S6). The shot displayed is confirmed to thereby change the length of the shot when necessary (S7). The division range is changed according to the method of changing the shot division range described referring to Fig. 4 and displayed as instructed (S8). The foregoing process is repeated until the videos of the shot agree with the image conceived by the user 17. The termination of the adjustment is instructed (S9), and the relevant shot is displayed on the display window 61 for displaying the post-adjustment shot (S10). The operation moves on to the next shot unless the operation is interrupted in response to the button 64-2 for ending the division or the adjustment up to a final frame is completed (S11). The completion of the adjustment may be followed by such a step as rearranging shots and incorporating the shot from other video files if necessary, and then the entire editing work is completed (S12).

[0024]

[EFFECT OF THE INVENTION]

As thus far described, the video edit supporting system

according to the present invention comprises an index supply section for recording index information serving as a marker for a content of the video and audio, a shot automatic division processing section for automatically detecting shots based on a plurality of references previously provided and showing, at the time of editing, a cut point defined based on another reference as a candidate with respect to an instruction to change a length of the shot defined based on a particular reference, and a display device of displaying a candidate shot. Therefore, because there is the interaction between the shot automatic detection and the user, the user only needs to browse the contents of the video file once so that the shot division is automatically performed with reference to the index information set by the user. The user can perform the editing work by only issuing the simple instruction to change the shot length automatically divided or rearranging the order of the videos per shot. More specifically, lengthy time consumed for confirming and recording the contents can be saved, thereby enabling the editing work to be efficiently performed, and it can be avoided that the videos are incorrectly edited due to the wrong memory and false recognition. Further, when the video contents cannot be confirmed beforehand, the speedy video editing can be achieved.

[0025]

The shot automatic division processing section employs the audio information as one of the plurality of references used

for the shot automatic detection. Therefore, the division is performed by means of the generation status of the particular speaker, thereby eliminating the occurrence of such a problem that the post-editing video contents may include any missing audio.

[BRIEF DESCRIPTION OF THE DRAWINGS]

[Fig. 1] Fig. 1 is a block diagram illustrating a configuration according to an embodiment of the present invention.

[Fig. 2] Fig. 2 is an illustration of a video division method according to the present invention.

[Fig. 3] Fig. 3 is an illustration of examples of a shot automatic division and an operation at the time of adjustment.

[Fig. 4] Fig. 4 is an illustration of an operation when a length of the shot division is changed according to the present invention.

[Fig. 5] Fig. 5 is an illustration of screens of display devices according to the present invention.

[Fig. 6] Fig. 6 is a flow chart of operation steps of the embodiment shown in Fig. 1.

[DESCRIPTION OF REFERENCE NUMERALS]

- 1 video source
- 12A reproducing/recording device
- 12B reproducing/recording device
- 13 main control section

- 13A shot automatic division processing section
- 13B index supply section
- 14 display control section
- 15A display device
- 15B display device
- 16 input unit
- 17 user

Fig. 1

- 1 display device
- 2 display control section
- 3 reproducing/recording device
- 4 main control section
- 5 shot automatic division processing section
- 6 index supply section
- 7 input unit
- 8 video source
- 9 user

Fig. 3

- 10 video file image
- 11 index information
- 12 pre-adjustment division information for display
- 13 shot division information based on respective references

Fig. 4

- 1 method
- 2 selection object
- 3 division point to be selected
- 4 next division point
- 5 current division point

Fig. 5

- 1 longer
- 2 longer in altogether
- 3 shorter

- 4 shorter in altogether
- 5 adjustment end
- 6 start division
- 7 end division
- 8 other file

Fig. 6

- 1 editing start
- 2 keyboard operation
- 3 no operation
- 4 operation
- 5 supply index information to relevant video frame
- 6 instruction to start division
- 7 memorize shot divisions based on a plurality of division

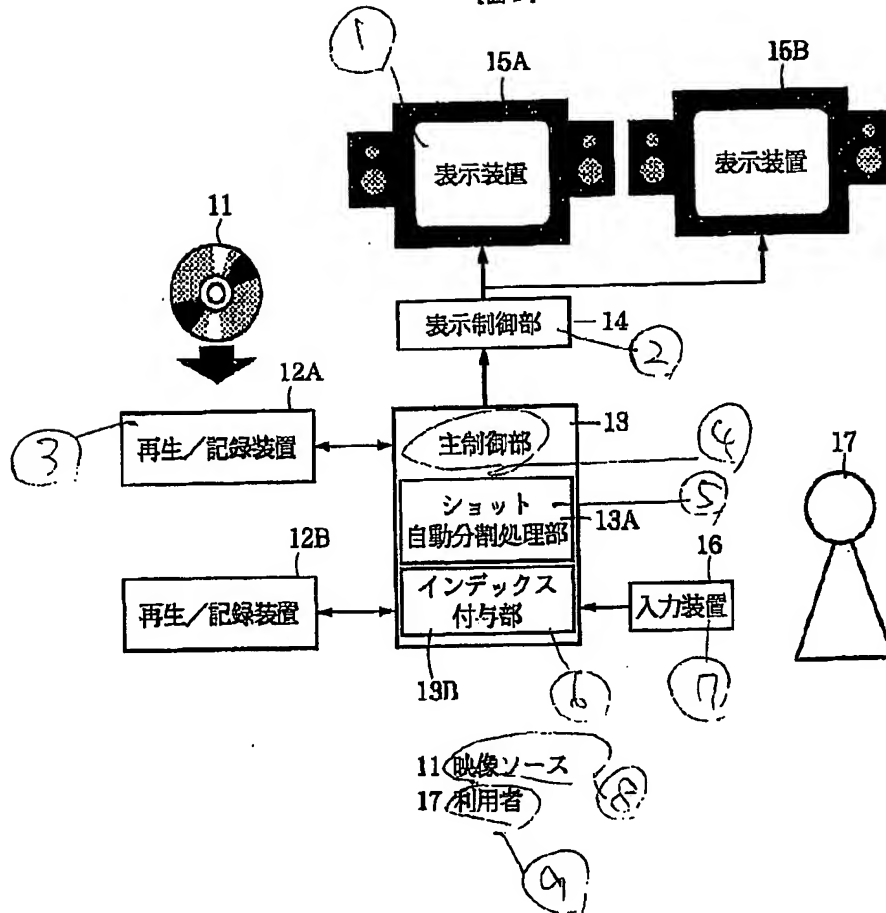
references

- 8 select and display pre-adjustment display shot
- 9 change shot length
- 10 YES
- 11 NO
- 12 change and display shot division range
- 13 instruction to terminate adjustment
- 14 change display window
- 15 interrupted or adjustment completed up to final frame
- 16 editing end

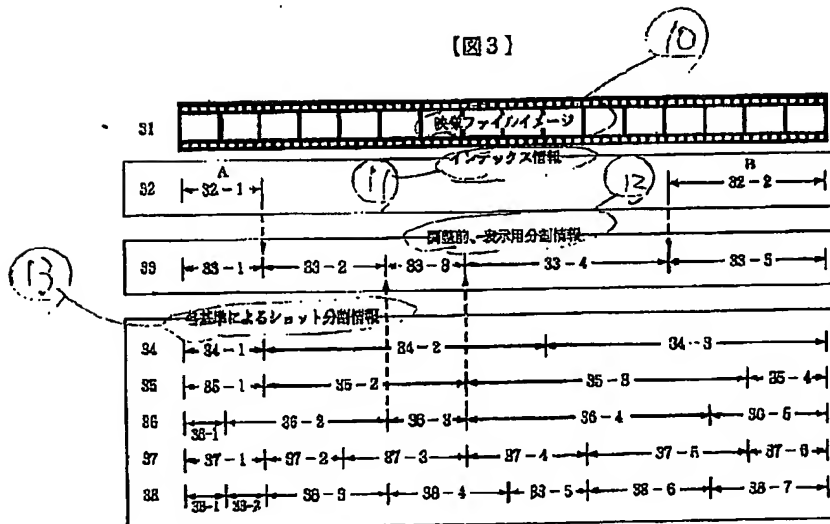
(6)

特開平5-290549

【図1】



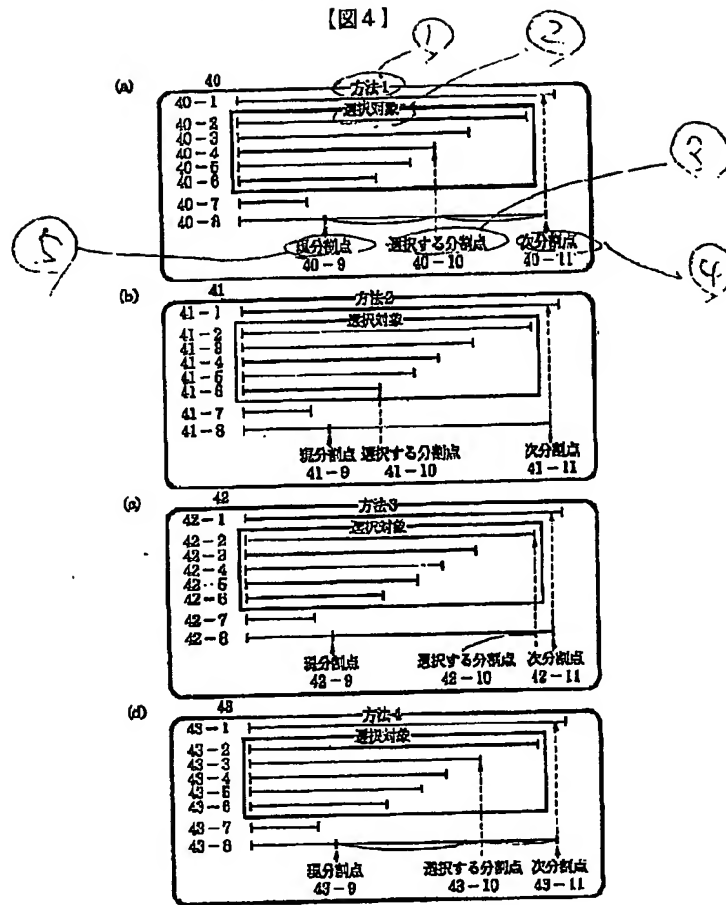
【図3】



(7)

特開平5-290549

【図4】

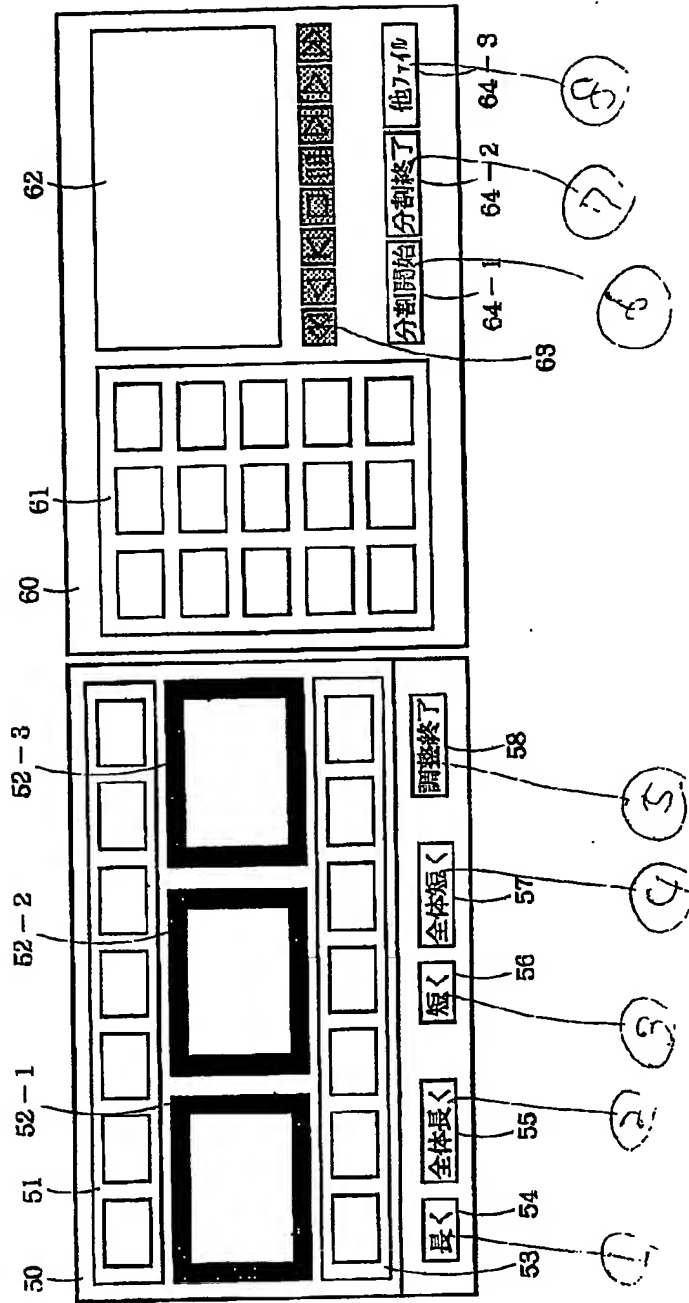




(8)

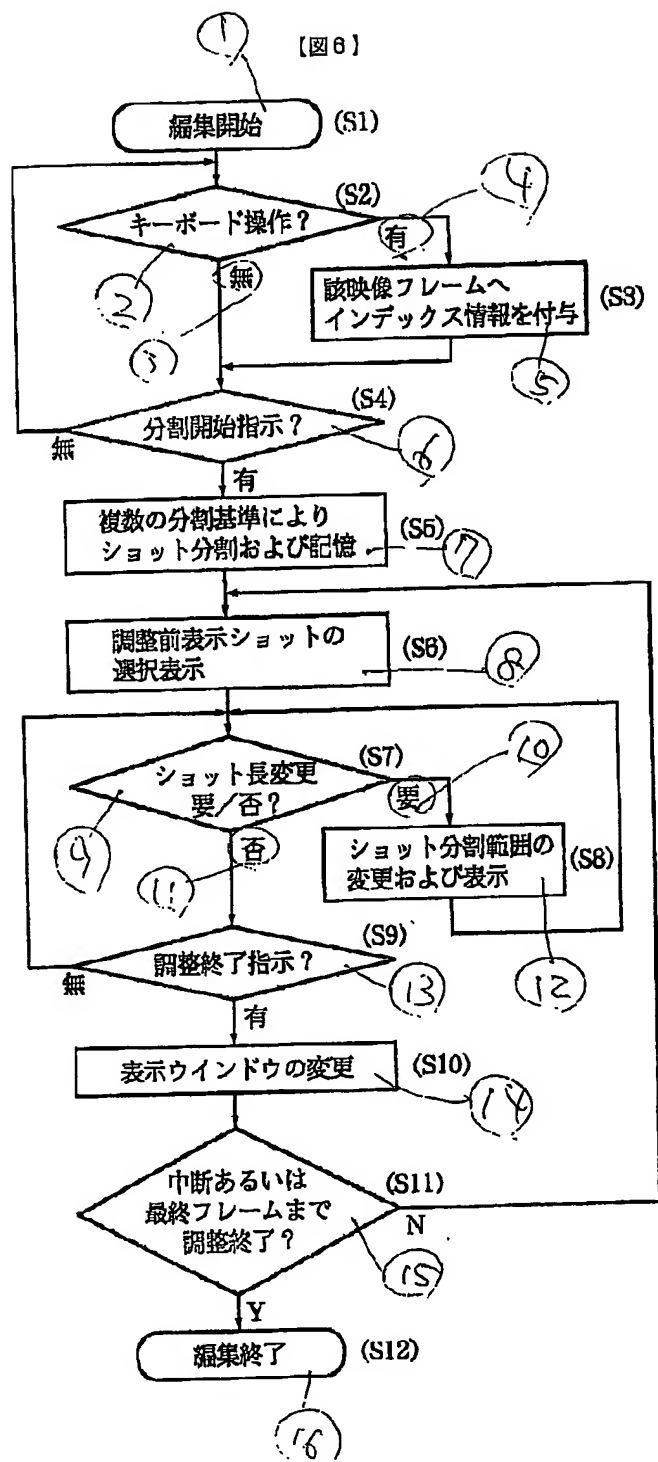
特開平5-280549

【図5】



(9)

特開平5-290549



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平5-290549

(43) 公開日 平成5年(1993)11月5日

(51) Int.Cl. <sup>5</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 1 1 B 27/034				
H 0 4 N 5/76	Z	7916-5C		
5/782	A	7916-5C		
		8224-5D		
			G 1 1 B 27/02	B

審査請求 未請求 請求項の数2(全9頁)

(21) 出願番号 特願平4-112404

(22) 出願日 平成4年(1992)4月6日

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区内幸町一丁目1番6号

(72) 発明者 長谷山 宏

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72) 発明者 大庭 有二

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72) 発明者 田中 知明

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

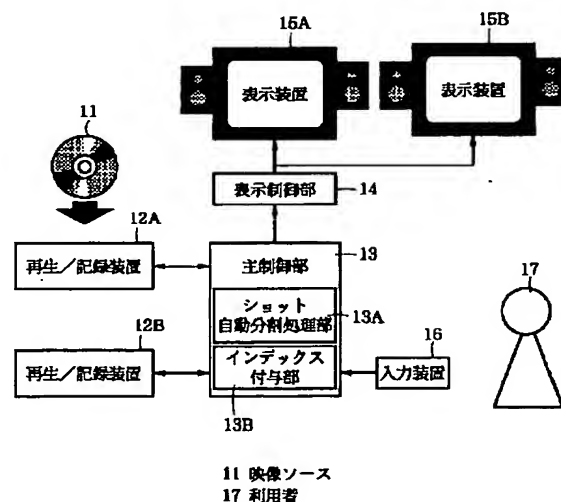
(74) 代理人 弁理士 小林 将高

(54) 【発明の名称】 映像編集支援装置

(57) 【要約】

【目的】 映像の編集に際し、複数の分割基準により複数のカットポイントを検出し、ショットの自動分割を行う。

【構成】 映像ソース11を再生/記録装置12Aで再生し、表示装置15Aに表示する。これを見ながら、操作者17は入力装置16から入力を行い、インデックス付与部13Bでインデックスを付与する。編集に際してショット自動分割処理部13Aは、あらかじめ設けられた複数の基準で分割された複数のカットポイントのうち、特定のものを選んで表示装置15Bに表示させる。そして、ショットの長さの変更指示を「長く」、「短く」などの簡単な指示で入力装置16から与えると、別のカットポイントが表示装置15Bに表示される。こうした処理を数回繰り返すだけで、映像編集を行うことができることを特徴としている。



1

## 【特許請求の範囲】

【請求項1】 映像および音声の編集を行う装置の支援装置であって、前記映像および音声内容に対する目印となるインデックス情報を記録するインデックス付与部と、あらかじめ設けた複数の基準によりショットを自動検出し、編集に際して、特定の基準により定められたショットの長さの変更の指示に対して別の基準により定められたカットポイントを候補として示すショット自動分割処理部と、候補のショットを表示する表示装置とを備えたことを特徴とする映像編集支援装置。

【請求項2】 ショット自動分割処理部は、ショットの自動検出に用いる複数の基準の1つとして音声情報を用いることを特徴とする請求項1に記載の映像編集支援装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、蓄積された映像の編集において、指示を与えると自動で映像を編集する映像編集支援装置に関するものである。更に詳しくいえば、一連の蓄積映像（以下、映像ファイルと呼ぶ）から、ある意味を持つ一連の映像（以下、ショットと呼ぶ）の自動分割を行い表示する。この表示を見て、利用者が「もっと長く」あるいは「もっと短く」といった簡単な指示を与えることにより、自動分割するショットの長さを変更し再度表示する。この繰り返しにより、編集作業を行う映像編集支援装置に関するものである。

## 【0002】

【従来の技術】 蓄積映像の編集において、映像ファイルの全てを注意深く視聴した後、映像内容のメモなどを頼りにショット分割を行っていた。また、静止画状態で映像内容を確認しながらショット分割する方法では、ショット連続位置の誤認などによって、当初のショット分割イメージと編集後イメージとの誤差が大きいのとなる。これを防ぐため、一連の動画をコマ送り状態で再生確認した後に必要な部分を複写するという作業を繰り返すことにより編集を行う場合など、編集後の映像イメージを確認するための、および音声のつながりを確認するために、確認の繰り返しが必要であり、迅速な編集や連続映像を見たときに考えた編集のイメージに対して、正確なショット分割が困難であるという欠点もあった。さらに、ショット分割に係わる全ての作業を自動化することはできず、手動で行うことしかできなかった。

【0003】 映像の場合には、検索のためのキーを付けにくいので、文字の場合のように容易には検索処理が行えない。ただ、動画の場合には、人間が目印になる信号、すなわち、インデックスを特定のフレームにあらかじめ付けておき、それを頼りに検索する方法などが使われている。この例として、ある種のビデオテープレコーダには、検索のための信号を録画開始時に対応づけてビデオテープに記録して、録画開始点の検索を効率化して

2

いるものや、レーザディスクにも同様な構成のチャプターサーチのような機能があるものがある。インデックスは、特定の映像の固まりに対しては機械的に付けることができる。しかし、録画開始位置等の単純な自動記録以上にインデックスを付けようとする、個々の静止画の内容を確かめながらの作業となるので、人手に頼る以外に方法はなかった。

【0004】 カットポイントの検出方法としては、長坂、田中：1990前期情処学全大1Q-5, pp, 642、大辻、外村、大庭：1991春信学全大D-675, pp, 7-387、大辻、外村、大庭：1991秋信学全大D-210, pp, 6-212などの方法がある。

【0005】 また、ショット分割基準の一例として、映像内の被写体に注目して考えると、物体領域の抽出方法としては、長尾、相馬、安藤他：1990信学技報IE90-105、阿居院、井手、長橋：1991秋信学全大D-267, pp, 6-269、乾、細野、山田：1991秋信学全大D-191, pp, 6-193などの方法がある。

【0006】 領域抽出後の処理としては、上田、間瀬、末永：1991秋信学全大D-220, pp, 6-222、井上、小山：1991秋信学全大D-200, pp, 6-202、小山、井上：1991TV学年次大ITEC'91 21-1, pp, 381などの方法がある。

## 【0007】

【発明が解決しようとする課題】 蓄積映像の編集において、ショット分割を行うためには、映像ファイルの全てを注意深く視聴する必要があり、時間がかかる作業である。また、静止画状態で映像内容を確認しながらショット分割を行った場合は、当初の連続表示時に考えた編集イメージと、編集後編集イメージとの誤差が大きいのとなってしまうという欠点があった。さらにこれを防ぐため、静止画による編集作業を行いながら、そこまででき上がった編集映像を確認するという作業の繰り返しを行わざるを得ず、大変に煩雑な作業であるという欠点もあった。

【0008】 画像の場合には、検索のためのキーを付けにくく、録画開始位置等の単純な自動記録以上にインデックスを付けようとする、人手に頼る以外の方法はなかったため、見た瞬間のイメージを、その映像フレームと対応して記録することは大変困難であった。

【0009】 本発明の目的は、カットポイントの検出に加え、ショットの自動分割も行いうる映像編集支援装置を提供することにある。

## 【0010】

【課題を解決するための手段】 本発明にかかる映像編集支援装置は、映像および音声内容に対する目印となるインデックス情報を記録するインデックス付与部と、あら

3

はじめ設けた複数の基準によりショットを自動検出し、編集に際して、特定の基準により定められたショットの長さの変更の指示に対して別の基準により定められたカットポイントを候補として示すショット自動分割処理部と、候補のショットを表示する表示装置とを備えたものである。そして、ショット自動分割処理部は、ショットの自動検出に用いる複数の基準の1つとして音声情報を用いたものである。

#### 【0011】

【作用】本発明においては、あらかじめ複数の分割基準により、短いショット長分割から長いショット長分割までの複数のショット分割を自動的にを行い、利用者がインデックスを設定してある部分は、同一インデックス情報が付けられている一連の映像フレームのショット分割に近い分割例を、インデックスが付けられていない部分は、分割長が短い分割と長い分割の中間の分割例を利用者へ提示する。利用者が自分の頭の中でイメージしているショット分割映像と比較して、「もっと長く」あるいは「もっと短く」といった簡単な対話を繰り返すことで、映像編集を行うことが可能となる。いいかえれば、20 複数の条件（基準）により映像の切れ目を自動検出するので、一度おおまかに映像ファイルの内容を見れば、そのとき使用者が設定したインデックス情報に基づいて表示されるショット単位の映像に対して簡単な指示を与え、ショットを並べる順番を決め、必要により複数の映像ファイルの内容と組み合わせるといった手順だけで編集を行うことが可能となり、映像フレーム番号とその映像内容の記録、利用者のイメージしたショット分割範囲の記録およびその映像内容のイメージ描写といった複雑な作業にかかわる無駄時間をなくし、効率的に編集作業を行える。さらに、記憶違い、誤認による誤編集をも回避できる。

【0012】そして、映像情報については、物体の領域抽出および領域抽出後の物体追跡などの処理に加え、領域抽出された各物体の特徴を抽出する。各物体の特徴抽出は、焦点深度、移動状況、変化状態、カメラ操作との関係などの条件により抽出を行い、ショット分割の基準とする。ショット分割の一例としては、抽出された1つ、あるいは複数の特定の物体が画面内にある間は1つのショットと見なす。また、シーンチェンジからシーンチェンジの間に挟まれた映像でも、特定の被写体が画面内になくなった場合は、ショットとして分割する場合もあり得る。

【0013】音声情報については、音楽の切れ目、音声認識による特定話者の発声状況、ノイズレベルやノイズ特性、などが分割基準となる。

#### 【0014】

【実施例】図1は本発明の一実施例を示す構成図である。図中の番号は次のものを表している。11は映像ソースで、レーザディスク、ビデオテープなどに記録され

4

た映像媒体である。映像ソース11は再生/記録装置12Aによって主制御部13へ取り込む。13Bは利用者17のキー操作によって映像内容にインデックスを付けるインデックス付与部である。利用者17が映像内容を見ながら任意のキーを操作することにより、映像フレーム番号とキーの内容を記録する。13Aはショット自動分割処理部であり、1つの映像ファイルに対して複数のショット分割の処理を行い記憶しておく。12Bは再生/記録装置であり、利用者17との対話により利用者17のイメージしているショット分割ができあがった時点で、20 ショット単位の映像を映像媒体に記録する。編集作業の最後にショット単位の映像の順番を並べ替える場合および他映像ファイルのショットを組み合わせる場合は、再生/記録装置12Aと12Bを交互に使用することで入れ替えを行う。14は表示制御部であり、ショット分割映像の表示および主制御部13の出力を表示するための映像合成制御部である。15Aおよび15Bは表示装置であり、表示制御部14により合成された内容を表示する。16はキーボードやマウスなどの入力装置であり、利用者17がインデックス情報やショット分割範囲の修正などを指示する装置である。

【0015】図2はカットポイント検出およびショット分割の例である。21-1~21-4はカットポイント検出の一例としてシーンチェンジ単位の分割を示した例である。すなわち、21-1、21-2、21-3、21-4の各々の間は、それぞれ撮影を一旦停止し、再度撮影を開始しているため、前映像内容との差分の大きいシーンチェンジ部分となっており、カットポイントの切れ目として検出している。22-1、22-2は利用者30 17が設定したインデックス情報による分割である。22-1は飛行機に着目し、22-2は自動車に着目していることが分かる。ただし、22-1と22-2の切れ目で、1枚の自動車の映像が飛行機の映像として誤設定されている。これは、通常速度の映像を見ながら設定したため、キーを押したり離したりするタイミングにずれがあることを示している。23は音声情報から分割した例であり、この映像は全て切れ目のない音楽が記録されている。このように、1つの映像ファイルでも分割方法により複数の分割が可能である。

【0016】図3はショットの自動分割例および調整時の動作を説明した図であり、31は映像ファイルを示す。32は前記利用者17の設定したインデックスで、この例ではインデックス文字A、Bが用いられている。インデックス文字A、Bは、例えば、飛行機と自動車を映像内容とする。33は調整前の表示用分割情報、34~38は複数のショット自動分割例を示す。はじめに、調整前の表示状況を見ると、インデックス32が設定されている部分はインデックス情報に基づいて分割し、インデックス32のない部分は中間的な分割を行っている分割36を採用している。

5

【0017】次に、利用者17の指示により分割点を変更する場合の処理を図4により説明する。ここでは、代表的な4方法について記述する。40~43は方法1~方法4である。40~8が図3の33に相当する調整前の表示対象となっている部分であり、利用者17の判断を要求している現分割点が40~9である。ここで、利用者から「長く」という指示があると、調整前の表示対象となっている分割方法の次の分割点40~11よりも時間的に前に存在し、現分割点よりも後に存在する。他の分割基準による複数の分割点、すなわち選択対象範囲より次候補を選択する。

【0018】方法1では、現分割点と次分割点の時間間隔を半分にした値に一番近い分割点40~10を選択し、この分割点の映像を利用者17へ表示し、次の指示を受ける。方法2では、選択対象の分割点の中で、時間的に一番早い分割点を選択する場合を示し、方法3では、一番遅い分割点を選択する場合を示している。また、方法4では、複数のデータから該当するデータを検索する場合に用いられる代表的な検索方法の式 $n/\epsilon$ を用いた分割点の選択を行っている。ここで、 $\epsilon$ は自然対数の底で、2.718、 $n$ は1または2である。

【0019】さらに、例は記述していないが、次分割点を越えた範囲を選択対象として選択したり、利用者17の分割点抽出傾向により、例えば、音楽の切れ目を分割点とする傾向があれば、それ以降の表示対象分割情報を当該、傾向にそった基準により分割したものと置き換える、などの制御を行うことにより、さらに、利用しやすい装置を構成できる。

【0020】図5は表示装置の画面構成例である。50と60は図1の表示装置15A、15Bに対応している。手順を追って説明すると、利用者17は、はじめに映像内容を一通り見る。これには、動画ウィンドウ62および映像操作ボタン63を使う。また、映像内容に対するインデックス情報を設定する場合は、図1に示したキーボードを使用する。自動分割の開始および終了は、ボタン64-1によ64-2により指示する。分割開始ボタン64-1を操作すると、各種の映像分割基準によりカットポイント検出およびショットの分割を開始し、複数の種類の分割を終了すると、自動分割されたショット単位の映像52-1~52-3を表示する。表示内容は映像52-1はショットの先頭の映像、映像52-2はショットの中間の映像、映像52-3はショットの最後の映像である。51は2通りの表示内容を表示する静止画ウィンドウであり、映像52-1の前、あるいは52-1と52-2の間が一番短く分割されているショットの頭の映像を表示する。同様に、ウィンドウ53は映像52-3の後、あるいは映像52-2と52-3の間の静止画を表示する。これにより、一度も映像内容を確認せずに利用することも可能である。

【0021】54、56に示した「長く」あるいは「短

6

く」等のボタンによりショット分割点の延長あるいは短縮を行う。延長あるいは短縮を行うショットの部分は、映像52-1あるいは52-3のどちらかの枠をクリックすることにより決定する。また、「全体長く」「全体短く」のボタン55、57が押された場合は、前後ともにショット分割点の延長あるいは短縮を行う。

【0022】ショット映像が利用者17のイメージに合致した時点で、「調整終了」のボタン58を押す。これにより、該ショットは調整後ショット表示ウィンドウ61へ設定される。ここでもショットの表示内容はショットの始め、中間および最後の映像である。このようにして、すべてのショットを調整し終わった場合、あるいは途中で分割終了ボタンにより調整を中断した場合、各ショット毎の編集に移る。ショット毎の編集方法としては、順番入れ替え、複写、削除、ボタン64-3による他の映像ファイルのショット組み込みなどが可能である。また、各ウィンドウに表示された映像にインデックスが付けられている場合は、映像合成表示することにより、利用者17の当初の分割イメージを確認できる。また、「長く」「短く」といった指示はマイクロホンによる音声指示としてもよく、これにより操作が極めて容易となる。

【0023】図6は主制御部の動作を説明したフローチャートである。なお、(S1)~(S12)は各ステップを示す。編集作業を開始し、はじめに(S1)利用者17が映像内容を確認する際、主制御部13ではキーの操作を監視(S2)している。もし、キーを押した場合は、該映像フレーム番号とキー内容をインデックス情報として記憶する(S3)。分割開始の指示(S4)がなければ引き続いてキーボード操作の監視を続ける。映像内容の確認が終了し、分割開始指示を行った場合、複数の分割基準によりショットの自動分割を行う処理および複数の分割範囲を記憶する(S5)。すべての種類の自動分割が終了した時点で、図3で説明した調整前の表示用分割情報33に従って表示を開始する(S6)。表示されたショットを確認して、必要によりショット長の変更を指示する(S7)。指示により、図4で説明したショット分割範囲の変更方法で、分割範囲の変更を行い表示する(S8)。この作業を利用者17のイメージに合うまで繰り返す。調整終了の指示(S9)により、該ショットをショット分割調整後のショット表示ウィンドウ61へ表示する(S10)。分割終了ボタン64-2による中断あるいは最終フレームまで調整が終わった場合以外は次のショットへ移る(S11)。また、調整が終了した場合には、必要によりショットの並べ替え、他の映像ファイルのショットの組み込みなどの編集を行い、編集作業は終了(S12)となる。

【0024】

【発明の効果】以上説明したように、本発明は、映像および音声内容に対する目印となるインデックス情報を記

7

録するインデックス付与部と、あらかじめ設けた複数の基準によりショットを自動検出し、編集に際して、特定の基準により定められたショットの長さの変更の指示に対して別の基準により定められたカットポイントを候補として示すショット自動分割処理部と、候補のショットを表示する表示装置とを備えたので、ショットの自動検出と利用者との対話によって、1回おおまかに映像ファイルの内容を見れば、そのとき利用者が設定したインデックス情報を参考として、ショット分割が自動的に行われる。利用者は、自動分割されたショット長に対して簡単な変更指示を与えたり、ショット単位の映像の順番を決めるなどの処理だけで映像編集を行うことが可能となる。すなわち、内容確認、内容記録にかかわる無駄時間を省き、効率的に編集作業を行える。また、記憶違いや誤認による不正確な映像編集を回避できる。さらに、映像内容を事前に確認することができない場合でも、迅速な映像編集を行うことが可能である。

【0025】また、ショット自動分割処理部は、ショットの自動検出に用いる複数の基準の1つとして音声情報を用いたので、特定話者の発生状況などを利用する分割により、編集後の映像内容で音声途中で欠落してしまうなどの不具合が発生しない。

【図面の簡単な説明】

8

【図1】本発明の一実施例の構成を示すブロック図である。

【図2】本発明の映像の分割方法の説明図である。

【図3】本発明のショットの自動分割例および調整時の動作を説明した図である。

【図4】本発明のショット分割長の変更を行う場合の動作を説明する図である。

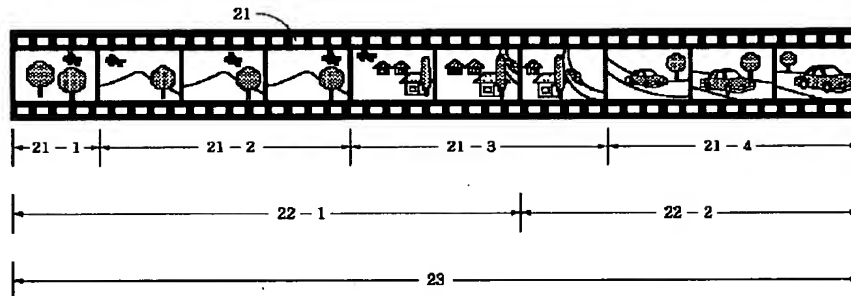
【図5】本発明の表示装置の画面構成の説明図である。

【図6】図1に示す実施例の利用手順を示すフローチャートである。

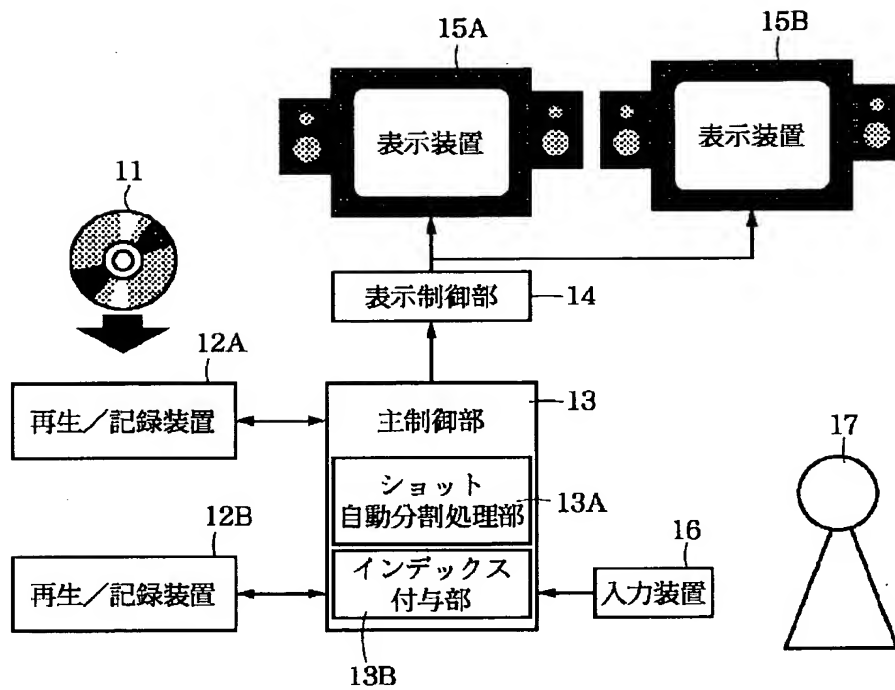
【符号の説明】

- 1 1 映像ソース
- 1 2 A 再生/記録装置
- 1 2 B 再生/記録装置
- 1 3 主制御部
- 1 3 A ショット自動分割処理部
- 1 3 B インデックス付与部
- 1 4 表示制御部
- 1 5 A 表示装置
- 1 5 B 表示装置
- 1 6 入力装置
- 1 7 利用者

【図2】

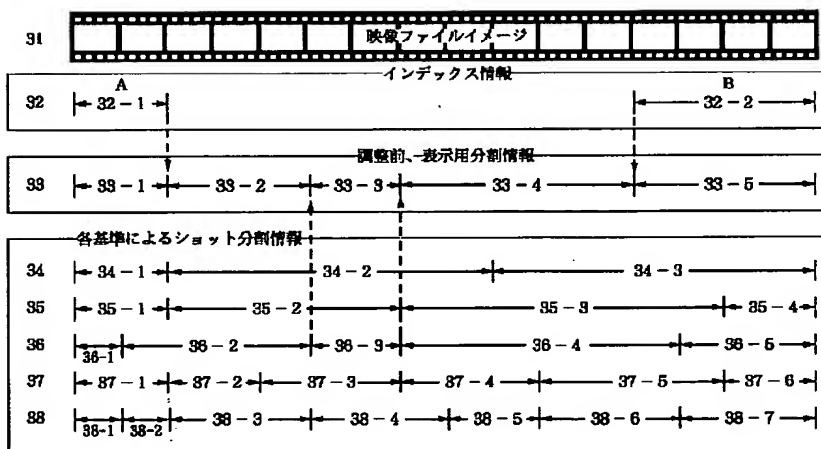


【図1】



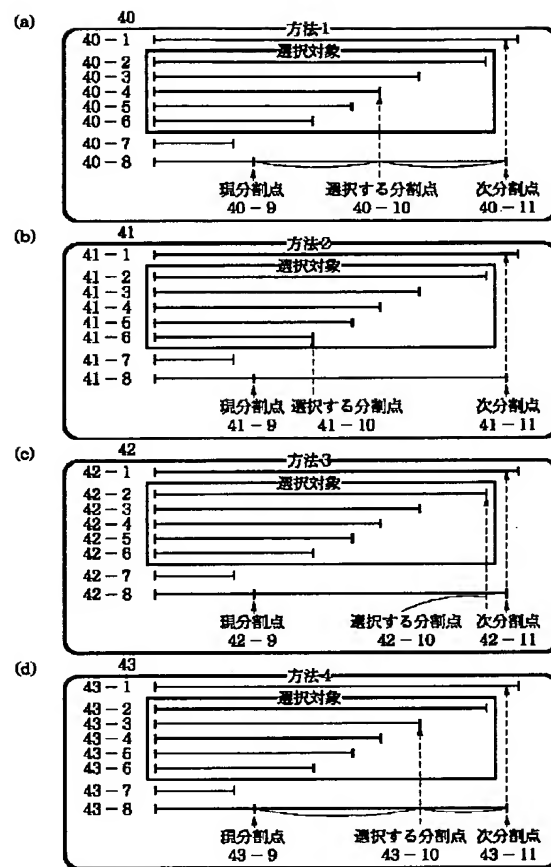
11 映像ソース  
17 利用者

【図3】

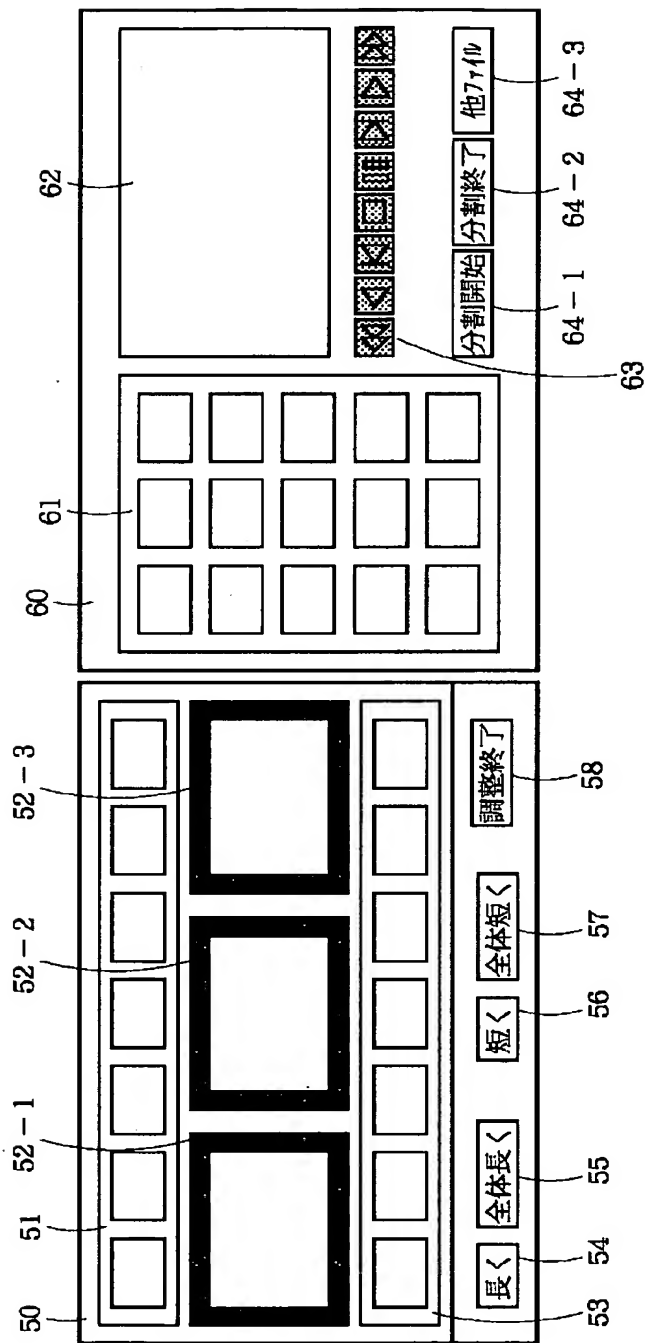




【図4】



【図5】



【図6】

